

第15章

統計的な分析と分析結果の解釈

英語教育実践は、複数の要因が複雑に絡み合う交互作用の中でその成果が定まる。実践のための指針を有効に得るためには、英語教育学研究の成果が適切に蓄積されなければならない。しかし、計算機やソフトウェアの急速な発展に伴い、統計的な分析に関する誤用や誤解釈も散見される。そのため、差の分析を解釈するという観点からは、効果量、検定力、信頼区間、そして図示といった手法を主軸に、研究成果をより有効に伝え、より有効に蓄積する方法を示した。また、複数の変数を使った分析と解釈という観点からは、相関分析、多変量分散分析、重回帰分析、探索的因子分析、そして共分散構造分析といった手法を取りあげ、多くの観測変数や、場合によっては潜在変数をも扱う多変量解析において、留意しなければならない点について指摘した。

キーワード

- 第1節 統計的仮説検定，探索的因子分析，重回帰分析，適性処遇交互作用
- 第2節 測定された差の分析，効果量，信頼区間，追試，データの可視化
- 第3節 複数の変数の分析，妥当性，信頼性，メタ分析

よいというだけではない。他者がそれを読めば同じことができるように詳しく、誤用や誤解釈がないように十分な情報を示しつつ、論文化することが必要である。そのため、本章では、本節でこれまで述べたように 1990 年代半ばからの時勢を振り返った後、次節以降において、差の分析と解釈や、多くの変数を用いた分析について、資料となる文献を引用しつつ、現状や、これから分析を行ったり先行研究を読むときに留意すべき点を指摘する。

〔前田啓朗〕

2 測定された差の分析と解釈

2.1 ARELE掲載論文に見る測定された差の分析

応用言語学系の国際ジャーナルでは、 t 検定や分散分析 (ANCOVA や MANOVA を含む) のような平均値に基づいた分析 (観測された差の分析) が大半であるということが報告されている (Gass, 2009; Lazaraton, 2005)。Plonsky (2013) は、1990 年から 2010 年までの 20 年間で *Language Learning* と *Studies in Second Language Acquisition* に掲載された論文のうち、統計的手法を用いた論文 606 本における分析手法の傾向と報告内容 (reporting practices) を調査した。その結果、半数以上 (56%) の論文で分散分析が使われていたという事実からも、グループ間の平均値差に焦点を当てた論文がいかに多いかということがわかる。

Mizumoto, Urano, and Maeda (2014) は、ARELE の第 1 号から第 24 号 (1990-2013) に掲載され、CiNii で公開されている論文 450 本のうち、76% (341 本) が量的研究であることを示した。国際ジャーナルと同じように量的研究がより多く掲載されている ARELE でも、観測された差の分析を行っている論文が多いことが予想できる。そのため、以下では、ARELE 第 13 号から第 24 号までの過去 12 号分における「測定された差の分析」の使用傾向を報告する。

ARELE 第 13 号から第 24 号までで、CiNii に公開されている論文 290 本のうち、統計的手法を用いているものは 255 本であった。そのうち、観測された差の分析を行っている論文が 186 本 (73%) であり、その他の 69 本 (27%) は相関係数を基にした分析 (重回帰分析やその他の多変量解析など) であった。この結果からも、ARELE に掲載されている論文の多くで、観測された差の分析が行われていることがわかる。

表 15.1 は観測された差の分析を行っている論文 186 本における分析方法と

報告内容を集計したものである。分析方法では、 t 検定のみが 30 本 (16%)、分散分析のみが 115 本 (62%)、その両方が 38 本 (20%) となっており、分散分析は合計で 186 本中 153 本 (82%) において使用されており、*ARELE* 掲載論文は分散分析の割合が高いといえる(対応するノンパラメトリック検定も集計に含む)。

「検定なし」の 2 本は、効果量とその信頼区間で解釈を行っているもの(Koizumi, 2011) と、記述統計と論旨に合った図示方法で議論を行っているもの(前田, 2008) の 2 本であるが、統計的検定の限界を理解した上で、必要な情報は十分に提示し、適切な解釈を加えているものなので、意図的に統計的検定を用いていない良い例であるといえる。「記載なし」の 1 本については、「有意な差 (significant difference) があった」とのみ記載されており、行っている分析を見ればおそらく分散分析であるということはわかったが、本文中に分析方法に関する記述は見られなかった。

表 15.1 観測された差の分析方法と報告内容 ($K = 186$)

	K	%
分析方法		
t 検定	30	16
分散分析	115	62
両方	38	20
検定なし	2	1
記載なし	1	1
報告内容		
サンプルサイズ	184	99
平均 (Mean)	174	94
標準偏差 (SD)	155	83
統計量	177	95
p 値 (*のみも含む)	182	98
p 値の明記	86	46
信頼性を示す指標	74	40
効果量	57	31
効果量の信頼区間	3	2

報告内容については、標準偏差 (standard deviation: SD) や信頼性を示す指標、 p 値の明記、そして効果量やその信頼区間が記載されていない論文が多く見られ、*ARELE* 掲載論文のうち、少なくとも、測定された差の分析を行っているものについては報告内容の改善が望まれるといえるだろう。

2.2 測定された差の分析と解釈における注意事項

高度な分析が可能な統計解析ソフトウェアの普及と、竹内・水本（2012）、平井（2012）、前田・山森（2004）などのように、英語教育学研究で用いられることの多い統計解析をわかりやすく説明している入門書が、ここ 10 年で発行されている状況が相まって、量的研究の中でも測定された差の分析が非常に多い現状であるが、前述のように報告内容については改善の余地がある。以下では特に報告内容に注目し、測定された差の分析と解釈における注意事項を挙げていく。（分析のための前提条件などの詳細は上記の入門書を参照されたい。）

(1) 再現に必要な情報を必ず書く

測定された差の分析を行っている論文において、標準偏差（*SD*）やその他の記述統計情報が報告されていないことがある。そのような場合、元データの再現や結果の検証が不可能である。そして、そのような論文は、量的研究において先行研究の効果量を統合して、その効果を検討するメタ分析で必要な効果量が計算できないために、メタ分析の対象から除外されやすい（Plonsky, 2013）。また、次節の「複数の変数を使った分析と解釈」にもあるように、変数間の相関係数も再現（二次分析）には必要になるため報告したい。さらに、*p* 値の報告についても正確な記述を心がけたい。*APA Publication Manual* 第 6 版（2010）では、（図表でのわかりやすさを重視する場合を除いて） $p < .05$ やアステリスク（*）での表記ではなく、 $p = .031$ というように明記することが求められており、0.001 以下の値を取る場合のみ $p < .001$ と表記する（p. 114）。

(2) 有意差至上主義からの脱却

統計的仮説検定の結果得られる *p* 値は、サンプルサイズに依存し、再現性に乏しい（Cumming, 2012）、そのため、水本・竹内（2008）は、実質的な差を検討するために効果量を用いる重要性を唱えており、*ARELE* 掲載論文でも 2011 年（第 22 号）以降は効果量の報告が増えてきている。 $p < .05$ やアステリスク（*）での表記では、「有意差あり・なし」の二分法のみ注目してしまい、多くの研究者・実践者が本当に関心のあるはずの「どの程度効果があるか」ということまではわからない。*ARELE* 掲載論文でも、報告すべき内容が書かれておらず、 $p < .05$ やアステリスク（*）のみの記載が優先されているものがあるという状況は、「有意差至上主義」のマインドセットのせいであるといえるだろう。

また、 p 値は再現性が低いため、再現性を重視するのであれば、追試(replication)によって結果を蓄積し、吟味していくメタ分析的アプローチを優先させるべきである。そのためには、効果量とその信頼区間を報告し、それに基づいた解釈を行うのが適切である(Cumming, 2012; 大久保・岡田, 2012)。ARELE掲載論文のうち、効果量と信頼区間による報告・解釈を行っている論文はKoizumi (2011)のみであるが、今後、そのような報告方法のさらなる普及が期待される。

(3) 実質的重要性 (practical significance) を考える

効果量や信頼区間をもってしても、実際に測定された差が、学習者や英語教育学研究にとって、どのような意味を持つのかという実質的重要性 (practical significance) はわからない。この解釈については、より多くの研究から得られた効果量とその信頼区間を用い、研究の文脈を理解した分析者自身が考えていく必要がある。

(4) 測定の信頼性を示す指標を報告する

ARELE掲載論文では、表 15.1 からわかるように、測定の信頼性を示す指標の報告が少ない (186 本中 40%)。いくら高度で複雑な分析を行っていたとしても、そもそもの測定の信頼性が低ければ、本当に差がある場合に有意差を検出できる力である検定力も低くなってしまい、結果の再現性が低くなる (検定力については、水本・竹内, 2011 を参照)。信頼性係数を使って各種の補正を行う二次分析もあることから、測定の信頼性を示す指標は報告すべきである。

(5) データの可視化を心がける

前田 (2008) に見られるような論旨に合った図示方法は、“A picture is worth a thousand p values” (Loftus, 1993) という言葉からもわかるように、 p 値のみでの結果提示よりも価値があるといえる (統計的検定は行っていないでも必要な情報があれば再現し検定もできる)。また、上位群と下位群に分けてから分析を行うような場合も、それぞれの群の分布がどのようなものであるかを見せるという目的から、生データが可視化できるような方法が好ましい。

(6) 要因は増やすよりも減らす努力を

一般的に、英語教育学のみならず心理学のような関連分野でも、先行研究の

結果から得られる検定力は低い。これはサンプルサイズが小さいことに起因するが、複雑なデザインで二元配置以上の分散分析を行う場合には、検定力も低くなりがちである。そのため、いたずらに要因を増やすよりも、先行研究をしっかりと概観することで、できるだけ少ない要因で仮説検証型の研究に落とし込む努力が必要であるだろう (Mizumoto, Urano, & Maeda, 2014)。

【水本 篤】

3 複数の変数を使った分析と解釈

3.1 ARELEの過去 10 年の動向

英語教育学研究には学習者・教員・学習環境などの多くの要因が複雑に関わる。そのため、研究目的に応じ、適切な分析手法を使うことが必要である。適切な分析手法を用いて初めて、より妥当なデータ解釈につながるため、研究のデザインを考える際に、どの分析手法を用いれば、どのような解釈ができるかを吟味する必要がある(水本, 2012)。

このような背景を受け、ARELE の収録論文でも複数の変数を対象にした分析が多く行われてきた。2004 年発行の第 15 号から現時点で最新号の 2013 年発行の第 24 号までの過去 10 年を概観すると、相関係数を用いた論文が 29 件(例 Ushiro, Hijikata, Shimizu, Nakagawa, Koga, Ohno, & Umehara, 2007; 山西・廣森, 2008), 多変量分散分析(MANOVA)を用いた論文が 2 件 (Adachi, 2013; Koga & Sato, 2013), 重回帰分析を用いた研究が 6 件 (例 Ohno, 2004; Sakuma, 2011), 探索的因子分析を用いた研究が 13 件(例 Fujita & Noro, 2009; Yamanaka & Takeuchi, 2009), 共分散構造分析(structural equation modeling)を用いた研究が 11 件(例 Katagiri, 2010; 斉田, 2012)ある。特に相関分析は基本的な分析手法であり、広く使われていることが分かる。また、これらの分析をすべて包括する共分散構造分析が、件数は少ないが継続的に使われている。適切な分析と解釈を行い、英語教育学研究をより良いものにするため、これら複数の変数を扱う分析手法を使う際に特に注意すべきことを 7 点述べる。

3.2 複数の変数を使った分析と解釈の注意事項

第 1 に、変数間の分析結果を、因果関係と直に解釈することはできない。因果関係と解釈できるのは、(a)原因が結果に時間的に先行していること、(b)原